

Finding the k -best Equivalence Classes of Bayesian Network Structures for Model Averaging

Supplemental Materials

Yetian Chen and Jin Tian
 Department of Computer Science
 Iowa State University
 Ames, IA 50011, USA
 {yetianc, jtian}@iastate.edu

Abstract

This technical report provides supplementary materials for the paper “Finding the k -best Equivalence Classes of Bayesian Network Structures for Model Averaging”. Section 1 provides the proof of proposition 1 in Section 2 of the paper. Section 2 gives the detailed algorithm for checking equivalence of two DAGs used by **Algorithm 1** in the paper. Section 3 presents more detailed experimental results.

Proofs of Propositions

Proposition 1 *For any decomposable score function that satisfies score equivalence, we have $score(G_W) = score(G'_W)$ if G_W and G'_W are equivalent over node set $W \subseteq V$.*

Proof. We can construct two equivalent DAGs G_V and G'_V over the total node set V from G_W and G'_W respectively. For each $v \in V \setminus W$, we pick arbitrary parent set $Pa_v \subseteq W$. Then G_V and G'_V can be constructed by $G_V = G_W \oplus_{v \in V \setminus W} Pa_v$ and $G'_V = G'_W \oplus_{v \in V \setminus W} Pa_v$. G_V and G'_V are equivalent since they have the same skeleton and same set of v -structures. This is because: G_W and G'_W have the same skeleton and the same set of v -structures; adding Pa_v 's for all $v \in V \setminus W$ only adds directed edges from any node $u \in W$ to node $v \in V \setminus W$. This produces the same skeleton as well as the same set of v -structures. By **Definition 3** and **Definition 4**,

$$\begin{aligned} score(G_V) &= \sum_{v \in W} score_v(Pa_v^{G_W}) + \sum_{v \in V \setminus W} score_v(Pa_v) \\ &= score(G_W) + \sum_{v \in V \setminus W} score_v(Pa_v), \end{aligned}$$

$$\begin{aligned} score(G'_V) &= \sum_{v \in W} score_v(Pa_v^{G'_W}) + \sum_{v \in V \setminus W} score_v(Pa_v) \\ &= score(G'_W) + \sum_{v \in V \setminus W} score_v(Pa_v), \end{aligned}$$

$$score(G_V) = score(G'_V).$$

Solving these equations yields $score(G_W) = score(G'_W)$. \square

Algorithms

$CheckEquivalence(G_W, G'_W)$ determines whether two DAG G_W, G'_W over W are equivalent.

Algorithm 1 $CheckEquivalence(G_W, G'_W)$

```

1: function CHECKVSTRUCTURE( $v, G_W, G'_W$ )
2:   for each pair of distinct  $u, w \in Pa_v^W$  in  $G_W$  do
3:     if  $u \notin Pa_w^W$  and  $w \notin Pa_u^W$  and
4:       ( $u \notin Pa_v'^W$  or  $w \notin Pa_v'^W$  in  $G'_W$ ) then
5:       return false
6:     end if
7:   end for
8:   for each pair of distinct  $u, w \in Pa_v'^W$  in  $G'_W$  do
9:     if  $u \notin Pa_w'^W$  and  $w \notin Pa_u'^W$  and
10:      ( $u \notin Pa_v^W$  or  $w \notin Pa_v^W$  in  $G_W$ ) then
11:      return false
12:    end if
13:   end for
14:   return true
15: end function
16: /* Check skeleton */
17: for each node  $v \in W$  do
18:   for each  $u \in Pa_v^W$  in  $G_W$  do
19:     if  $u \notin Pa_v'^W$  and  $v \notin Pa_u'^W$  in  $G'_W$  then
20:       return false
21:     end if
22:   end for
23:   for each  $u \in Pa_v'^W$  in  $G'_W$  do
24:     if  $u \notin Pa_v^W$  and  $v \notin Pa_u^W$  in  $G_W$  then
25:       return false
26:     end if
27:   end for
28: end for
29: /* Check  $v$ -structures */
30: for each node  $v \in W$  do
31:   if CHECKVSTRUCTURE( $(v, G_W, G'_W)$ )=false then
32:     return false
33:   end if
34: end for
35: return true

```

Experiments

In this section we provide more experiments to evaluate the ability of using our method for structural discovery.

In first experiment, we randomly generated three synthetic networks over 15 variables, and drew 100, 200 and 500 samples from the joint distributions they represent. We then computed the posteriors of all 210 possible edges by model averaging over the corresponding k best ECs or DAGs ($k = 1, 10, 100$). For comparison, we also computed the exact posteriors by full model averaging over all possible networks using algorithm from (Tian and He 2009). We plot the ROC curves in Figure S1. Similar to the findings in (Tian, He and Ram 2010), the accuracy for the structural discovery evaluated by the area under ROC (AUC) is a non-decreasing function of k . For same k , the accuracy for model averaging over the k -best ECs is significantly better than that over the k -best DAGs.

Another observation concerns about the reliability of using MAP model (structure) for structural inference. In the second plot, a best DAG (AUC = 0.57) is a bad model for inferring the structure as it is only slightly better than a random structure. Using the best EC achieves much better estimation.

However, using only the best EC is not necessarily a good idea either as showed in the Figure S2. In this case we study the Tic data set. In Table 1 of the main paper, we observed that the DAGs in the top 10 ECs have the same posterior probability. We compare the structure of these 10 ECs, represented by CPDAG, in Figure S2(a). It shows these 10 structures only agree on the edge $9 \rightarrow 4$. Other edges (even the reversible edges) vary among these CPDAGs. Since these 10 ECs are equally probable, and the number of DAGs in each of them is almost equal, using any of these ECs will result in unreliable inference in the network structure.

For further comparison, we present the network structure obtained by averaging over all possible DAGs (by computing the exact posterior probabilities of all edges) in Figure S2(b), and the network structure obtained by averaging over the DAGs in the top 10 ECs in Figure S2(c). In these networks, we only include the most probable edges, those with posterior probabilities greater than 0.4. We set this threshold such that the edges above this threshold do not form any cycles in the structure. The corresponding edge probabilities are also showed in the figure. We first observe that the two structures (and edge probabilities) are very close to each other, differing only in the existences of edges $7 \rightarrow 9$ and $0 \rightarrow 9$ and the direction of edge $8 \rightarrow 1$. This means even using a small number of ECs ($k = 10$) may give an estimation close to full model averaging. In Figure S2(b), except edge $9 \rightarrow 4$, all other edges have probability of only 0.5 or lower, indicating the high uncertainty on the network structure. This is consistent with the observation in Figure S2(a) that there are a great number of distinct yet equally probable structures. Finally, the edge probabilities provide certain confidence information on the corresponding edges in the network.

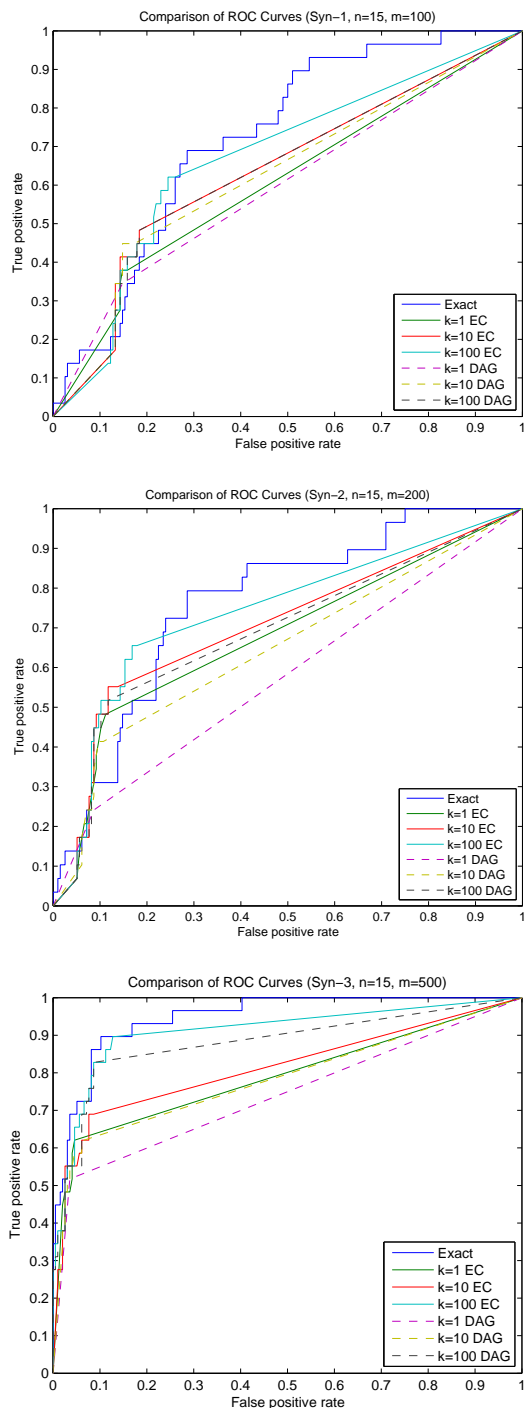


Figure S1: Comparison of ROC curves for edge discovery in Bayesian networks

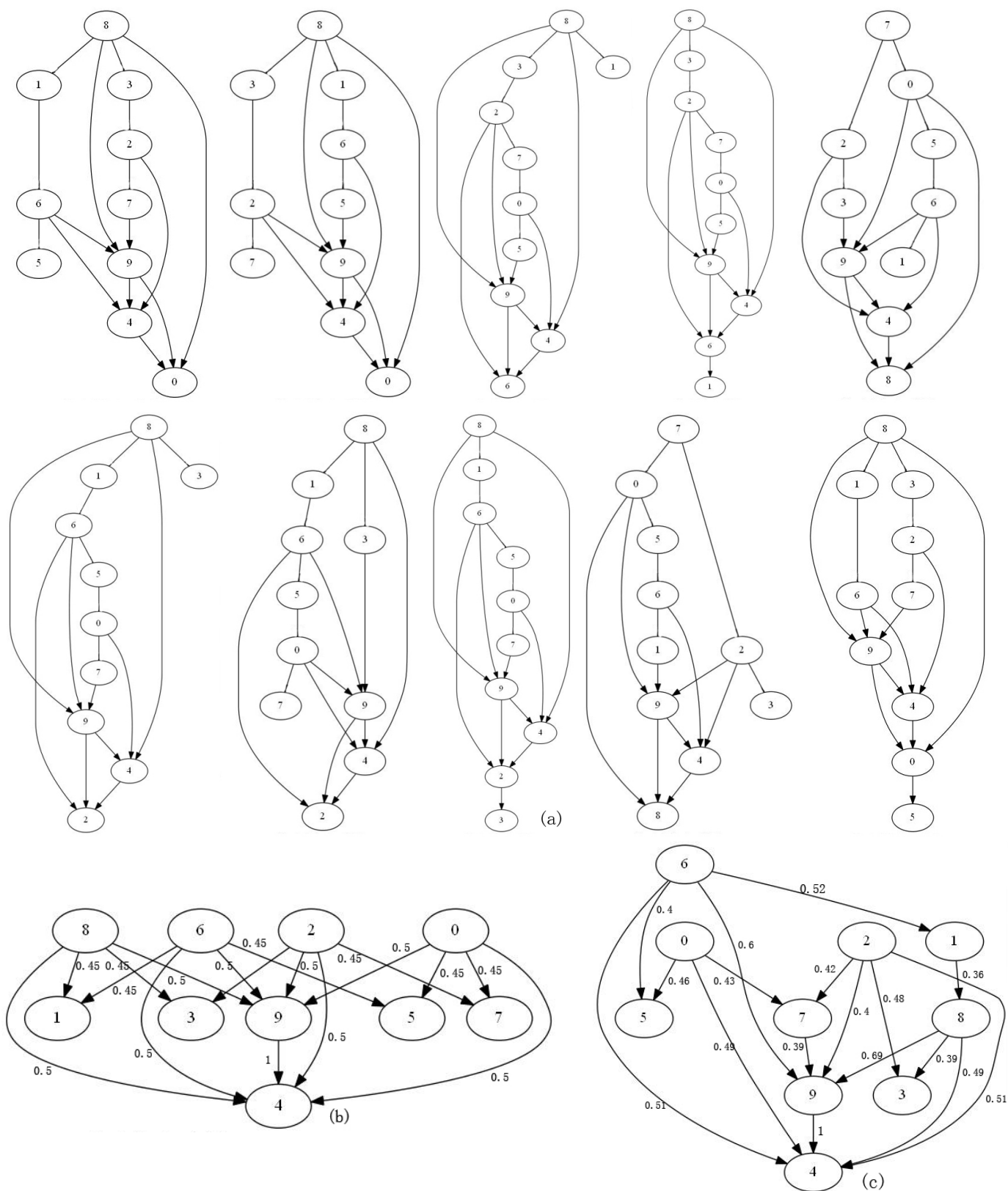


Figure S2: Results for Tic data set. (a) The top 10 equally probable equivalence classes for Tic-Tac-Toe Data set. Each equivalence class is represented by a CPDAG where reversible edges are depicted as undirected edges, compelled edges are depicted as directed edges. (b) Network structure averaging over all DAGs. (c) Network structure averaging over top 10 ECs. The numbers besides the edges in (b) and (c) indicate the estimated posterior probabilities of edges. In these networks, we only include the most probable edges, i.e., edges whose posterior probabilities are greater than 0.4. We set this threshold such that the edges above this threshold do not form any directed cycles in the structure.